GRADIENT0

Sicheres Machine Learning auf sensiblen Daten

Mit der DQ0 Datenquarantäne

2019

www.gradient0.com

Daten und Maschinelles Lernen

Der rasante Aufstieg der Künstlichen Intelligenz und insbesondere ihrer wichtigsten Disziplin, dem Maschinellen Lernen, basiert auf drei Faktoren: wissenschaftliche Durchbrüche in der Informatik, ausreichend hohe Rechenkapazität und Daten. Viele Daten. Daten sind die Grundlage der KI-Renaissance, die längst viel mehr ist als ein Trend. Datengetriebene Prozesse sind so erfolgreich, dass sie den Fortschritt über Industrien hinweg maßgeblich mitbestimmen und auf lange Sicht mitbestimmen werden. Daten sinnvoll und in großem Maßstab zu nutzen, heißt Wachstum und Profitabilität stärken und neue Möglichkeiten zu erschließen, die ohne KI und ohne die Daten nicht denkbar wären.

Dabei treten zwei wichtigen Fragen auf:

- Wie kann ich meine Daten für Maschinelles Lernen bereitstellen?
- Wie kann ich garantieren, dass ich Eigentümer meiner Daten bleibe und die Verarbeitung der Daten vollständig sicher und kontrollierbar bleibt?

Um datengetriebene Prozesse zu entwickeln oder zu implementieren, benötigt es die Hilfe von Experten aus dem Bereich „Data Science“, das sind speziell ausgebildete Informatiker oder Mathematiker, die erweiterte statistische Analytik und Methoden des Maschinellen Lernens beherrschen. Als Unternehmer habe ich zwei Möglichkeiten: entweder ich baue innerhalb meiner Organisation selbst ein oder mehrere Data Science Teams auf oder ich hole mir diese Leistungen über Partner. Für beide Wege brauche ich eine sichere Schnittstelle, die definiert, wie welche Daten dem internen oder externen Data Science Team zur Verfügung gestellt werden.

Oft handelt es sich bei den in Frage kommenden Daten um solche, die schützenswerte Informationen enthalten. Datenschutz bekommt, nicht zuletzt aufgrund der Datenschutz-Richtlinie der EU und den entsprechenden nationalen Verordnungen, endlich die Aufmerksamkeit, die er verdient. „Data Ownership“ ist ein wichtiges Thema. Ich muss sicherstellen, dass ich zu jedem Zeitpunkt die Kontrolle über meine Daten behalte. Personenbezogene Daten oder Firmengeheimnisse dürfen niemals (ohne explizite Einwilligung) preisgegeben werden.

KI wird erwachsen

Maschinelles Lernen ist keine neue Disziplin mehr. Ihre Werkzeuge und Methoden werden immer ausgereifter und einer breiteren Gruppe von Fachleuten zugänglich. Gleichzeitig etablieren sich Geschäftsprozesse zu Daten-Marktplätzen, auf denen Unternehmen ihre Daten auf schnelle und einfache Weise Datenwissenschaftlern zur Verfügung stellen, um davon durch neue Lösungen oder optimierte Verfahren selbst zu profitieren oder direkt den Wert der Daten zu realisieren.

Künstliche Intelligenz ist überall. Auch wenn ihre Methodik noch immer ganz am Anfang steht, sind ihre Anwendungen und Werkzeuge für bestimmte Domänen bereits sehr ausgereift. Oft werden allgemeine, vortrainierte Modelle des Maschinellen Lernens verwendet, im Falle der Sprachverarbeitung zum Beispiel solche, die auf einem großen, allgemeinen Wortschatz entwickelt wurden, und dann auf den vorliegenden Daten „nachtrainiert“, also spezialisiert für den einzelnen Anwendungsfall. Aber wie sieht die Schnittstelle für dieses Verfahren aus?

Der Standardweg heute: die Daten werden dorthin gebracht, wo sie von Datenwissenschaftlern verarbeitet werden können. Um den Datenschutz zu gewährleisten werden die Daten vorher üblicherweise pseudonymisiert (Ersetzung von personenbezogenen Entitäten), generalisiert oder verrauscht (so verallgemeinert oder verändert, dass einzelne Datensätze nicht mehr erkennbar sind) oder minimiert (vorgeblich nicht gebrauchte Eigenschaften der Daten entfernt).

Dieser Ansatz ist keine Lösung. Die Bearbeitung der Daten zu ihrem Schutz ist aufwändig und domänspezifisch, es gibt keine allgemeingültige Lösung hierfür. Für manche Daten (z.B. bestimmte Texte oder Bilder) kommt eine Anonymisierung kaum in Frage. Außerdem ist dieser Prozess fehleranfällig und nicht ausreichend sicher. In zahlreichen Publikationen und Wettbewerben wurden vermeintlich ausreichend anonymisierten

Daten geheime Informationen entlockt. Es gibt keine Methode, um algorithmisch zu bestimmen, ob Daten selbst tatsächlich erfolgreich unkenntlich gemacht wurden.
Und wie, genauer: wo, werden die Daten eigentlich pseudonymisiert? Auch hier stellt sich die Frage nach einem sicheren Zugang zu den Daten – ein Henne-Ei Problem.

Hinzu kommt, dass dieser Ansatz das Problem der Schnittstelle zu den Werkzeugen des Maschinellen Lernens nicht löst. Das jeweilige Data Science Team muss diese selbst mitbringen bzw. integrieren.

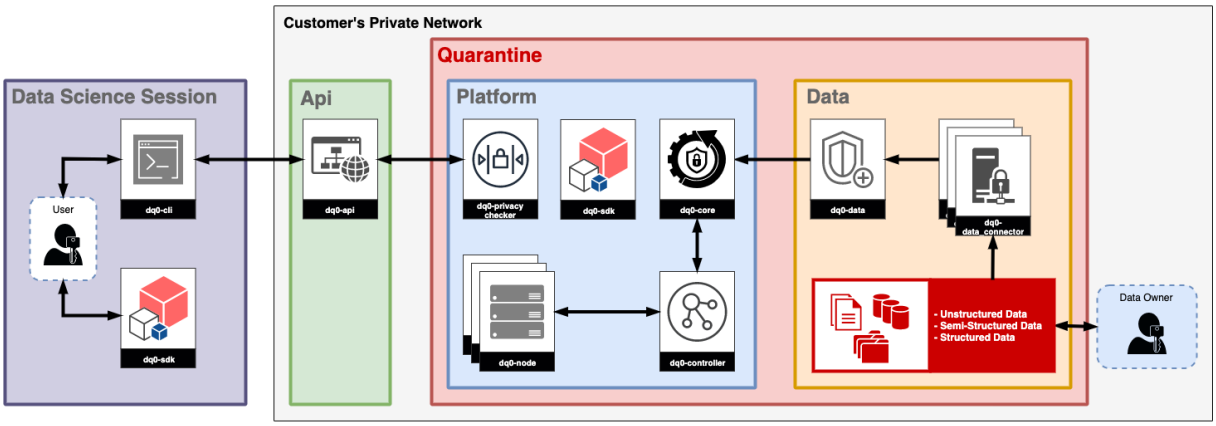
DQ0

Die DQ0 Datenquarantäne löst beide Probleme. Sie bietet einen sicheren Hafen für die Daten und macht sie gleichzeitig derart für die Analyse zugänglich, dass ein Data Science Team mit ihren gewohnten Werkzeugen direkt mit den Daten arbeiten kann.

DQ0 dreht den Spieß um: die Daten kommen nicht zur Analyse, die Analyse kommt zu den Daten. DQ0 schließt die vorhandenen Daten in einer Quarantäne ein. So bleiben sie sicher geschützt vor unbefugtem Zugriff. Die Daten werden zu keinem Zeitpunkt verändert oder herausgegeben.

DQ0 implementiert den Datenschutz zum Zeitpunkt der Anfrage an die Daten. Zugrundeliegend ist hier das Prinzip der „Differential Privacy“, welches Gradient Zero auf den Bereich des Maschinellen Lernens ausgeweitet hat. Jede Antwort, welche die Quarantäne über die sichere DQ0 API verlässt, wird auf strenge Sicherheits- und Datenschutz-Konformität überprüft.

Die DQ0 API bietet damit eine sichere Schnittstelle, um Maschinelles Lernen auf den Daten durchzuführen. Das Data Science Team kann iterativ KI-Modelle entwickeln, ohne dass sensible Informationen über die Daten preisgegeben werden.



Wie im obigen Diagramm zu sehen, wird DQ0 vollständig innerhalb des Netzwerkes des Kunden installiert. In der optionalen Cloud-Variante von DQ0 wird das Netzwerk ebenfalls so eingerichtet, dass nur der Kunde Zugriff und Steuerungsmöglichkeit hat. Lediglich der Port für die DQ0 API bleibt nach außen geöffnet. Jede Antwort, welche die Quarantäne verlässt, wird vorher vom „DQ0 Privacy Checker“ überprüft.

Der „Data Owner“ hat die vollständige Datenhoheit. DQ0 kann dabei mit beliebigen Datentypen umgehen: strukturierte Daten genauso wie Textdokumente, Bilder oder beliebige andere Formate.

Der Zugriff auf die Daten erfolgt read-only mittels der DQ0 Datenkonnektoren und des DQ0 Data Shield.

Nutzer, welche Modelle auf den Daten entwickeln möchten, melden sich über die DQ0 API an und verwenden das DQ0 SDK, um die Modelle zu definieren und zu testen. Über die API und den DQ0 Command Line Client

können die Modelle dann an die DQ0 Kunden-Instanz gesendet werden. Dort werden sie auf den echten Daten trainiert. Rückgabewert dabei ist eine vom DQ0 Privacy Checker freigegebene Antwort über die Güte der aktuellen Modellversion.

Die DQ0 API ermöglicht zudem allgemeine, aggregierte Informationen und Schemata der Daten zu erfragen, sowie synthetische Daten zu erzeugen, um die erforderliche Daten-Vorverarbeitung und das Feature Engineering zu erleichtern.

Über die intelligente DQ0 Schnittstelle können auf einfache Weise Standard-Modelle in die DQ0-Instanz übertragen werden, um ihre Brauchbarkeit für das konkrete Problem schnell zu testen.

Mit der optionalen studi0 Integration, welche in der Cloud-Version von DQ0 erhältlich ist, kann zudem eine umfassende Data Science Plattform für Workflow Management, Daten- und Modell-Versionierung und hoch-paralleles Modell-Fitting genutzt werden.

Differential Privacy und Maschinelles Lernen

DQ0 versteckt die Daten des Kunden konsequent in einem sicheren, inneren Netzwerk. Da die Modelle zu den Daten kommen und nicht umgekehrt, müssen keinerlei Informationen über die Daten leichtfertig herausgegeben werden. Wie aber stellt DQ0 sicher, dass die Antworten, welche die externen Datenwissenschaftler erhalten, ausreichend geschützt und gleichzeitig für ihre Arbeit brauchbar sind?

DQ0 setzt auf das Prinzip der Differential Privacy, welches in der Quarantäne in allen Phasen der Machine Learning Entwicklung konsequent zum Einsatz kommt.

Differential Privacy (DP) ist eine Definition, welche mathematisch garantiert, dass jeder, der das Ergebnis einer DP-Analyse sieht, dieselbe Schlussfolgerung über die privaten Informationen einer Person zieht, unabhängig davon, ob die privaten Informationen dieser Person in die Eingabe für die Analyse einbezogen wurden oder nicht.

Anders ausgedrückt: Anfragen an einen Datensatz unterscheiden sich nach DP nicht (oder nur in einem definierbaren, beschränkten Maß), wenn ein Datenpunkt dem Datensatz hinzugefügt oder jener aus diesem entfernt wird. Daher lassen sich bei DP-konformen Abfragen keine Rückschlüsse auf einzelne Datenpunkte (z.B. Personen) ziehen.

Differential Privacy garantiert Schutz vor:

- Membership disclosure (z.B. hat eine Person einen bestimmten Account?)
- Attribute disclosure (die Aufdeckung einer Eigenschaft eines Datenpunktes)
- Identity disclosure (die eindeutige Zuordnung eines Eintrages zu einer Person)

DP ist mathematisch definiert als: Eine randomisierte Berechnung *M* erfüllt die ε-Differential Privacy, wenn für alle benachbarten Datensätze *x* und *x'* und jede Teilmenge *C* der möglichen Ergebnisse Menge *(M)*, gilt:

$$Pr[M(x) \in C] \leq \exp(\epsilon) \times Pr[M(x') \in C]$$

Das heißt, das Verhältnis der Wahrscheinlichkeiten, dass zwei benachbarte Datensätze die gleiche Antwort geben, ist durch den Faktor ε (bzw. $exp(\varepsilon)$) begrenzt.

Nun erfordert jedoch der Schutz von Informationen nach diesem Prinzip im Bereich des Maschinellen Lernens besonderer Sorgfalt. Denn mit erweiterten statistischen Methoden ließe sich diese DP-Eigenschaft ausnutzen, um über ausreichend viele Anfragen Rückschlüsse auf einzelne Datenpunkte zu erzielen.

Ein möglicher Angriff könnte ein sogenanntes „shadow model“ darstellen, das seinerseits die DP-konformen Ausgaben auswertet, um zu schützenswerten Informationen zu gelangen. Die einfachste Methode ist hierbei die Ausnutzung des „overfitting gap“, um ein Membership disclosure zu erreichen. Dieser Unterschied zwischen den Fehlerkurven von Training und Test ist größer je spezifischer das Modell für die Testdaten trainiert wurde. Da Modelle des Maschinellen Lernens ihre Qualität dadurch erlangen, dass sie sich im Trainingslauf anhand von präsentierten Trainigsdaten so anpassen, dass sie die Summe der Fehler der Differenzen der

vorhergesagten Werte und der Zielwerte minimieren und so „von den Daten lernen“, kann es leicht passieren, dass sie, ausgestattet mit genügend Freiheitsgraden die Daten einfach auswendig lernen. Ihre Performanz auf den präsentierten Trainingsdaten ist dann sehr hoch, die auf neuen Testdaten aber gering. Der genannte Angriff macht sich diese Eigenschaft zu Nutze: er betrachtet die Konfidenz der Vorhersage eines präsentierten Datenpunktes (also wie sicher ist sich das Modell, dass der Datensatz z.B. in eine vorhergesagte Kategorie fällt) und schließt aus einer hohen Konfidenz, dass dieser Datenpunkt Teil des Trainingsdatensatzes gewesen sein muss und somit im eigentlich versteckten Datensatz vorkommt.

Der DQ0 Model Checker begegnet diesem Angriff mit einer automatischen Prüfung jedes Modells, welches innerhalb der Quarantäne ausgeführt werden soll. Nur wenn diese Prüfung entlang der implementierten DP Kriterien bestanden wird, darf das Modell über die DQ0 API über das eigentliche Training hinaus von außen verwendet werden.
Eine Anwendung des Modells innerhalb der Domäne des Kunden ist jederzeit möglich.

Die DQ0 Datenquarantäne stellt zu jedem Zeitpunkt sicher, dass die API nur solche Informationen über die Daten nach draußen gibt, die auch mit erweiterten statistischen Verfahren niemals zu einem Bruch des Datenschutzes führen. Sie setzt dazu auf innovative Verfahren zur Datensicherheit für Maschinelles Lernen.

Die Methode ist wissenschaftlich überprüft; die Implementierung wird vom TÜV Austria zertifiziert.

Tl;dr

Daten sind wertvoll. Sie können die Grundlage bilden für die Optimierung bestehender Prozesse, die Entwicklung neuer Einsichten und Möglichkeiten und für anhaltenden wirtschaftlichen Erfolg. Dies kann jedoch nachhaltig nur gelingen, wenn die Sicherheit der Daten zur obersten Priorität erhoben wird. Dafür hat Gradient Zero die DQ0 Datenquarantäne entwickelt – von Datenwissenschaftlern für Datenwissenschaftler, kompromisslos gebaut für Datensicherheit. Mit DQ0 erhalten interne oder externe Data Science Teams Zugang zu sensiblen Daten, ohne dass sensible Informationen preisgegeben werden.

DQ0 ist die sichere Schnittstelle von Daten zu Machine Learning. Verschlüsselung und datensichere Anfragen werden bei DQ0 nicht einfach durch definierte Regeln eingefordert, sondern sind als Prozesse algorithmisch integraler Bestandteil der DQ0 Software. Das DQ0 SDK bietet eine standardisierte Schnittstelle für allgemein verfügbare genauso wie für selbst entwickelte Modelle.

Ob sensible medizinische Daten, personenbezogene Nutzerinformationen oder geheime Maschinendaten, DQ0 ist die einfache und sichere Schnittstelle in die datengetriebene Zukunft.

Über Gradient Zero

Gradient Zero ist ein führendes Unternehmen für maschinelles Lernen mit Sitz in Wien, Österreich. Wir bieten branchenübergreifende Lösungen für maschinelles Lernen.

Künstliche Intelligenz hat sich zu einem etablierten Geschäftsfeld und Forschungsgebiet entwickelt. Erfolgreiche KI-Projekte erfordern jedoch gut konzipierte Lösungen mit sorgfältig ausgewählten Techniken und einem umfassenden Verständnis der zugrunde liegenden Methoden, ihrer Implementierungsanforderungen und den Ansprüchen an die Datensicherheit. KI ist hier, um die Dinge einfacher, schneller und komfortabler zu machen - das ist unsere Mission.

Kontakt

e-mail: office@gradient0.com
phone: +43 660 4259199
web: www.gradient0.com

GRADIENT0